Developed for the

California Environmental Technology Certification Program

Measurement & Monitoring Protocol

California Environmental Protection Agency
September 1998

PERFORMANCE-BASED CERTIFICATION OF HAZARDOUS WASTE MEASUREMENT AND MONITORING TECHNOLOGIES

GUIDANCE TO TECHNOLOGY PROPONENTS

Prepared by

Ruth R. Chang and G. Wolfgang Fuhs

California Environmental Protection Agency
Department of Toxic Substances Control
Hazardous Materials Laboratory

For additional copies, download from the internet at www.calepa.ca.gov/calcert/protocols

or contact
Office of Environmental Technology
(916) 327-5789
(916) 445-6011 FAX
oet@arb.ca.gov

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PROGRAM BACKGROUND

Assembly Bill 2060 (Weggeland, 1993, now Section 25200.1.5 of the California Health and Safety Code) authorized the Department of Toxic Substances Control within Cal/EPA (hereafter referred to as "DTSC" or "Department") to establish a certification program for hazardous waste-related technologies which include technologies for pollution prevention, waste treatment, measurement and monitoring. The objectives of this program are promoting the development of innovative technologies, simplifying the permitting process, eliminating redundant verifications, and accelerating the diffusion of new environmental technologies in the U.S. and abroad. Subsequently, the certification programs within Cal/EPA were expanded to certification by other Cal/EPA regulatory bodies, Air Resources Board and Water Resource Control Board. The DTSC certification program is supported by the U.S. EPA as a pilot program for Environmental Technology Verification. In addition it enjoys the support and endorsement of other U.S. government agencies. This guidance addresses the elements that are required for the verification of *measurement and monitoring technologies including those used in site characterization.*

INTRODUCTION

Certification of California environmental technologies is based on a technology's verified performance; verification extends to performance claims set forth by the technology owner (in the following referred to as *"proponent"* or *"manufacturer"*). These claims apply to a technology's performance under certain operating conditions as represented to the user. Performance-based certification by itself does not provide regulatory relief, but establishes a strong basis and encouragement for users and jurisdictions, including but not restricted to California state environmental agencies, to accept the technology for their respective programs, provided that the certified performance meets their needs. To enable prospective users to make this determination, the Department issues an evaluation report which elaborates upon the technology's specifications and its demonstrated performance and on the conditions under which the technology can be expected to perform as claimed. Certification of a technology for a specified range of applications does not preclude other uses, except that such uses are solely a matter between the manufacturer and the user. The term *"verification"* refers to the technical process of independently and successfully authenticating a technology's performance. The term *"certification"* refers to an administrative act by the State of California which requires, and is based on the verification of a technology.

Performance-based certifications are different from other certifications which evaluate technologies against fixed protocols or regulatory standards. In performance-based certification, quality assurance criteria are applied to validate performance claims to ascertain the performance of the technology for a specified range of applications. In addition to efficacy, the California certification program evaluates the reliability, protectiveness, and manufacturing of a technology. Protectiveness refers to protection of the public health, the operator of the technology, and the environment. Manufacturing refers to the production, distribution, and use of a technology in a quality-assured manner.

Within Cal/EPA's Department of Toxic Substances Control, the program office for technology certification is the Office of Pollution Prevention and Technology Development in Sacramento, California. The lead office for protectiveness evaluations is the Human and Ecological Risk Division in Sacramento. The lead office for measurement and monitoring technology evaluations is the Hazardous Materials Laboratory in Berkeley, California. The Department is committed to protect the intellectual property rights of the proponent to the extent provided by law. The term of certification is normally three years; recertification is required to maintain certification status. During the period of certification, the manufacturer is held to maintain the quality of the technology at a level equal to or higher than the level at which it was certified. If a significant improvement of the technology occurs, the proponent may elect to have the technology recertified after reevaluation and for any newly established claims and/or conditions of use.

The Department may, however, develop and apply minimum performance standards and thereby exclude technologies which do not contribute towards Cal/EPA's goals of public health protection and environmental quality improvement. To be accepted into the process, a technology must also meet criteria of statutory eligibility and commercial readiness, and is subject to program-related factors. A flowchart presenting a typical process for the certification of a Hazardous Waste Environmental Technology is included in this document for information. The technology verification process may vary slightly among departments and boards within Cal/EPA.

This protocol provides a description of documents and performance data that are required for the evaluation of measurement and monitoring technologies. Measurement and monitoring technologies include those used in site characterization, environmental field testing, sampling, sample preparation methods, and analysis by instrumental, chemical and biological methods. The performance standards established for this certification program are mostly consistent with Guidance for Methods Development and Methods Validation for the RCRA Program (Ref. 1) and other federal programs for environmental technology verification. The material presented in this test protocol is mainly for technologies used for laboratory and field monitoring and measurements. Some of the data requirements in this protocol are not applicable to all measurement and monitoring technologies (such as geophysical technologies). In all cases, the data requirements will be individually discussed with the applicant to define the scope of work prior to making a certification services agreement.

INFORMATION NEEDS AND EVALUATION PRINCIPLES

The materials presented by the proponent should describe and explain the scientific and engineering basis of the technology. They should also reflect the studies undertaken by the proponent to develop and buttress the performance claims presented to the Department. These studies are collectively referred to as "*in-house validation studies.*" The extent and quality of these studies not only reflects the proponent's efforts in establishing the performance claims but also assists in determining the scope and focus of the *verification study* which is to follow. In addition, in-house validation studies lay the basis for evaluating the proponent's quality management system.

In addition to in-house validation studies there may be *"third-party validation studies."* These may have been commissioned by the technology proponent or by some other party. Third-party validations, if they are independent, quality-assured, and in fact verify the performance claims put forth by the proponent, may in part or fully take the place of the *independent validation studies* that are otherwise required to establish certifiable performance.

In principle, the same quality standards apply to in-house validations, third-party validations, and verification studies. Verification studies are designed to produce quality-assured and independent data *only to the extent that they are not yet available but are necessary to substantiate the proponent's performance claims and satisfy other legitimate considerations* (such as reliability, protectiveness, applicability and manufacturing).

To initiate the process leading to certification, the technology proponent needs to supply the Department with certain documents. The documentation consists of the following elements: (1) description of the technology; (2) proof of eligibility; (3) proof of concept; (4) performance claims and scope of applications; (5) proof of attributes through data validation; (6) data analysis and acceptance criteria; (7) applications; (8) reliability; (9) quality control process in manufacturing; (10) safety issues associated with use; and (11) training requirements. These elements are discussed in detail in the remainder of this guidance. A checklist for data package and information is presented at the back of this document.

1 Description of the Technology

A summary statement should give the name of the technology and proponent's ownership or operational control over the technology. The summary should include a short description of the underlying scientific and operating principles, the range of applications related to measurement and monitoring of hazardous wastes, potential applications to hazardous waste treatment, pollution prevention, or to regulatory compliance. Limitations of the technology should also be disclosed.

2 Proof of Eligibility

In the context of this program, *"technology"* means a system which includes equipment and/or materials, knowledge and operating skills as reflected in user manuals or training, and the practice of quality assurance in manufacture and application. Professional services which apply professional skills to develop individual solutions to individual problems are not certifiable under this program, although certified technologies as defined here can be components of a custom-engineered system. DTSC will evaluate each technology as a complete operating system with respect to its efficacy, reliability, and protectiveness. Eligibility for certification under the DTSC statute requires that the system presented is a "technology" as defined above.

Technologies accepted under this program must be hazardous waste-related. Incineration technologies are specifically excluded from certification by California law. The technology must be either in channels of commerce or ready for commercialization, and that the information and data described below are available or can be procured in a reasonably short time period (within one year) for evaluation. This does not include possible delays caused by local permit requirements and logistic constraints related to field demonstrations. DTSC attempts to assist proponents to find suitable field sites or coordinate field studies for cost-effective technology demonstrations.

A technology should provide environmental and/or economic advantages over existing and generally available technologies. Hazardous waste-related measurement technologies find application in site characterizations, site remediation, effluent monitoring, and monitoring of process streams. The Department is especially interested in innovative technologies for measurements in the field, including both "screening" and "quantitative" analytical measurements. Measurement technologies that are not exclusively hazardous-waste related may be evaluated jointly with other Cal/EPA jurisdictions for possible multi-media certification. Environmental technologies not related to hazardous waste would be referred to other Cal/EPA certifying bodies.

Depending on workload, the Department will attempt to schedule evaluations in a mutually agreeable manner. It may retain the qualified professional services of verification entities who are bound to confidentiality and free from conflict of interest. Proponents will be informed in advance of any such plans and will be given the opportunity to express concern or give their approval.

3 Proof of Concept

The technology must be based on sound scientific and engineering principles and appropriate standards. The scientific basis of a technology can be based on the following documents:

a. Principles generally accepted in the pertinent field(s) of science and engineering;
b. Recognition in textbooks and the peer-reviewed literature;
c. Research and data that are of a quality acceptable to experts in the field.

Since certification involves an extrapolation from a limited knowledge base on a technology's performance to a range of applications, DTSC staff should be in a position to understand the scientific basis of the technology to anticipate its strengths and to design verification studies. Therefore, DTSC staff critically reviews the submitted data and conducts literature searches and consultations as necessary while protecting proprietary data to the extent provided by law.

4 Performance Claim(s) and Scope of Applications

Based on the validation studies conducted in the laboratory and/or pilot scale studies in the field, the proponent will have developed performance claims for the technology. Typically, successful in-house validation is the incentive for the proponent to manufacture the technology, and the performance claims are used in marketing. For these reasons, the Department considers the proponent's performance claims the hypotheses that are subject to testing in the process of verification.

Performance claims must be clearly expressed. Each claim must be meaningful, measurable, verifiable, and relevant to environmental protection. Performance claims for measurement technologies must be clearly defined with respect to the target analyte(s), concentration ranges, the target environmental medium or media, and operating conditions. Here "media" refers to different types of sample, such as water, soil, air, etc. It is advised

to formulate performance claims on media and matrices for which there are sufficient data and information to support the performance claims. Performance claims may relate to efficacy, reliability, and protectiveness. "Efficiency" as a reference to savings of time and cost from the use of the technology clearly is another desirable property, especially because the Department will not consider certification of technologies that are inefficient.

It should be noted that the performance claims initially advanced by the proponent as hypotheses for testing may not be those that are ultimately adopted for certification. If performance is better than expected, performance claims may be tightened. When performance claims are relaxed to reflect proven performance, however, the reduced level of utility of the method needs to be weighed before verification is considered a success.

5 Proof of Attributes through Data Validation

In the following, specific data needs and treatment of data are discussed separately for in-house validations and field verification studies.

5.1 In-House Validation Study

In-house validation studies are performed by the technology developer or manufacturer and usually have been documented as part of *technology development record.* The proponent is expected to generate sufficient data to explain and buttress the performance claims. Data requirements for the *in-house* validation are technology-specific. The proponent should submit *data packages* that are relevant to the operation of the technology. In general, the following elements are included.

5.1.1 Standard Operating Procedures

5.1.1.1 Sampling procedures, if they are a part of the technology, must be established for the media or matrices of concern (water, air, soil, etc.); these include description and operation of the sampling device, sample preservatives (if necessary), etc.

5.1.1.2 Sample preparation procedures specific for each type of media or matrices, such as soil, sediment, sludge, water, etc.

5.1.1.3 Standard analytical procedures for the analysis of target analytes, a detailed description of analytical procedures, including apparatus, equipment, chemicals, preparation of reagents, etc. required for the analysis.

5.1.1.4 Environmental factors and other variables that affect the accuracy of the results; including pH, temperature, moisture (of soils), reagent volumes, reaction times, sample and reagent storage conditions, etc.

5.1.1.5 Instrument calibration: acceptance criteria and QA/QC requirements.

5.1.1.6 The dynamic range of effective quantitation.

5.1.1.7 Utility requirements, such as electrical power supply, cooling system, ... , etc. for the installation of the equipment, etc.

5.1.2 Sensitivity

Data for evaluating the sensitivity of a chemical test typically include method detection limit(s) (MDL), quantitation limit (QL) for the target analyte(s) and the probability of false results relative to a target concentration under specified operating conditions.

5.1.2.1 Method Detection Limit

The method of determining the detection limit must be stated. In the U.S. EPA RCRA program, the MDL is defined as the minimum concentration of a substance that can be measured and reported with 99% confidence that the analyte concentration is greater than zero. The MDL is determined from analyses of a sample containing the analyte in the matrix of concern (Ref.2, Appendix A). The MDL is calculated from multiplying the appropriate one-side 99% t-statistic by the standard deviation (*s*) obtained from a minimum of four analyses of a matrix spike containing the analyte of interest at a concentration 3 to 5 times the signal-to-noise ratio. The term "matrix" refers to the diversity of physical and chemical properties within one environmental medium (soil, water, air, solid waste, etc). For example, water matrices may include groundwater, surface water, leachate, wastewater, etc. Soil matrices may include clean or contaminated sand, loam, clay, etc. The standard deviation should be determined from a minimum of seven replicate analyses of each analyte in a relevant environmental matrix.

Equation 1 $MDL = t_{(n-1,\ 1-\alpha = 0.99)} \times s$

where
t = critical value obtained for the Student's t distribution
n = the number of analyses
α = the error rate
s = standard deviation of n data points

5.1.2.2 Quantitation Limit

For a quantitative method, the quantitation limit is the lowest analyte concentration that can be determined in the sample with 95% confidence level. A calibration curve is established by analyzing the analyte of interest at 5 concentration levels in triplicates. The response factor of each data point must be within ±10% of the mean value. To reflect the matrix effect, the appropriate sample extraction and dilution factor(s) multiplied by the lowest acceptable calibration standard is used to determine the method quantitation limit (QL).

Equation 2 $QL = LS \times F$

where:
LS = the lowest concentration standard used for the five point calibration curve
F = the sample extract dilution factor times any additional factors to account for matrix interferences

Sensitivity must be demonstrated in at least three different matrices in each medium. From a practical sense, the matrices selected for demonstration should be those on which the technology is commonly used, not the matrix for which the method is most sensitive. The dynamic concentration range for effective quantitation should also be determined. In addition to the original data, results indicating MDL, QL and dynamic range tested should be summarized as indicated in Table 1.

Table 1 Method performance data on MDL, QL, and Dynamic Range

Sample type	Number of samples	*MDL*[1]	*QL*[2]	*Dynamic Range Tested*
Matrix 1	(≥ 7)			
Matrix 2	(≥ 7)			
Matrix 3	(≥ 7)			

[1] Analyses of one blank and a minimum of seven replicates of each matrix spiked with the analyte of interest at concentrations 3 to 5 times the signal-to-noise ratio, MDL = t (n-1, 1-∝ = 0.99) *x* s.

[2] QL = the lowest acceptable calibration standard concentration as determined by a 5 point actual curve fitting × the sample extract dilution factor × any matrix factors.

For a semiquantitive method, depending on the design of the technology, generally the sample concentration may be expressed in two classes: (1) analyte(s) above or below a target concentration; and (2) analyte(s) within a concentration range. The detection limit of a semiquantitative method is defined as the lowest target concentration or concentration range that can be identified by the technology at the 95% confidence level under specified operating conditions.

To avoid excessive misidentification, a screening method should have a detection limit 10 times below the action level, and a quantitative method should have a quantitative limit at least 5 times below the regulatory action level. The target concentration can be project-specific according to the intended use of the data, or the regulatory level enacted by the federal or California Code of Regulations for the classification of hazardous wastes (Ref. 3, 4), or the suggested risk-based level for ecological system (Ref. 5, 6).

5.1.2.3 Rates of False-Positive and False-Negative Results

The testing should demonstrate the probability of false positive (FP) and false negative (FN) results in detecting the substance(s) of interest at a specified target concentration(s) as described above. A semiquantitative ("screening") method should be able to detect the analyte (or analytes) of interest or below relevant regulatory action levels. The probability of false-positive and false-negative results should be determined under the specified operating conditions in a matrix of concern. A false-negative response is a negative result for a sample that contains the analyte(s) above the target level or the level declared in the performance claim. A false-positive response is defined accordingly as a response above the target level for a sample that contains analytes below that level.

For *in-house* validation, a minimum of 20 blank samples spiked at one-half the target or the action level may be analyzed to determine the rates of false positives and a minimum of 20 blank samples spiked at two times the target or the action level to determine the rates of false negatives (Ref. 7).

The interpretation of results obtained relative to an action or regulatory level varies with the precision and accuracy of the method and the prevailing data quality objectives of the study. The rates of false positive and false negative results should be reported at a specified confidence level. For example: at the 95% confidence level the rates of false-negative results would be less than 5% and the rates of false positive results would be less than 20%.

5.1.3 Interferences

It is necessary to determine the effect of potentially interfering compounds. This includes: (a) compounds that are expected to coexist on a site with the target analyte(s); (b) compounds that are chemically related to the target analyte(s); and (c) other compounds that may elicit the same response and are sufficiently common or environmentally relevant. Separate studies should be conducted to establish the effect of interferences by organic and inorganic compounds. Negative interferences, such as the effect of masking agents, should also be investigated whenever they are suspected to exist. Twenty samples that contain a 50- to 100-fold excess of known or potentially interfering compounds should be analyzed to evaluate their effect on the performance of the method. A scheme for the summary of the false positive, false-negative and interferences is given in Table 2.

Table 2 Summary of False Positives, False Negatives, and Interferences Studies

Sample types	Number of samples	Number of samples less than the action level[4]	Number of samples greater than the action level[4]
Sample spiked at ½ the action level[1]	20		
Sample spiked at 2x the action level[2]	20		
Interferences[3]	20		

[1] Spiked samples for the determination of the rate of false-positive results.
[2] Spiked samples for the determination of the rate of false-negative results.
[3] Clean Matrix or environmental samples contain a 50- to 100-fold excess of interfering compounds or compounds potentially present.
[4] The regulatory levels, project-specific DQO, or the suggested risk-base level.

5.1.4 Specificity and Cross Reactivity

The specificity of the technology for the target analyte (or analytes) needs to be determined. This applies especially to bioassays, such as immunoassay, enzymatic analysis, or other biosensors, etc., that respond to several analytes. For testing methods which give a signal response to multi-analytes, such as polychlorinated biphenyls (PCBs), polyaromatic hydrocarbons (PAHs), volatile organic compounds (VOCs), etc., the relative responses of components must be determined. Concentrations of other analytes that elicit the same response as a given concentration of the target analyte (or analyte used in calibration) may be listed in a table to express their equivalent concentrations. In immunoassays it has become customary to state the minimum concentration of an analyte that gives a false positive result at 50% of the response relative to the target compound. The response to mixtures of these components needs to be investigated to determine interactions, synergetic or antagonistic effects, between these components. This information is essential for data interpretation. For example, most BTEX immunoassay test kits commonly use xylene as a calibrator. The signal response to benzene, the analyte of health concern, is lower than that of either toluene, ethyl benzene, or the xylene isomers. Therefore, if benzene is the analyte of concern, confirmation by a qualitative and quantitative laboratory analysis is necessary. To correctly use these analytical data, the user may either apply a conversion factor or recalibrate the testing system with the mixture of analytes that is actually present at the site.

5.1.5 Recovery Data and Matrix Effects

The technology should be suitable for one or several environmental matrices. Performance data should be obtained for a minimum of three matrices. The term of "matrix" is provided under section 5.1.2.1. The selection of matrices for the test should be in accordance with the proposed applications. Depending on the intended use of a technology, in some cases samples taken from different sources or geographic areas may be used instead of samples representing three different matrices. For example, if a technology is proposed mainly for groundwater monitoring, water samples taken from different geographic areas or from different sources are acceptable for the method validation.

For determinations of method accuracy and precision, either standard reference materials or spiked matrices containing known amounts of target analyte(s) are used. If the method includes extraction of the sample, target analytes should be directly spiked into the real-world samples prior to extraction, not into the sample extracts. If feasible, the spiked matrix should be aged under environmental conditions for one to two weeks to mimic real-world samples before sample extraction takes place. Spiking of extracts is used only if the technology uses the same extraction procedure as the reference method and the validation is for the measurement component of the technology only.

For the purpose of verification study a sufficient quantity of homogeneous sample should be prepared and split for replicate analyses. For bulk materials, the sample should be thoroughly homogenized and mixed to avoid sampling errors. Split samples must be analyzed both by the test method and by the reference method. Matrix-specific performance data should be obtained by analyzing replicate aliquots of spiked matrices at the suggested concentration levels. A matrix blank should be concurrently analyzed with each batch of samples to assess background interference. Statistical analysis and acceptance criteria for the evaluation of method accuracy and precision are described in the following section.

5.1.5.1 Accuracy

Accuracy refers to the difference between the sample result and the true concentration in the sample. Bias refers to the consistent deviation of measured values from the true values. Bias can be caused by the extraction process, matrix effects, interferences, and contamination within or from the analytical system. The accuracy of the method is assessed by the percent of recovery (R) through replicate analyses of proficiency evaluation samples (PES), or spiked matrices containing known amounts of target analytes, or the Standard Reference Materials (SRMs). Method performance can be established by analyzing ten replicate aliquots of three different matrices spiked at low and high concentrations. These analyses can also be performed on the contaminated environmental samples or on the matrix spikes. The suggested low concentrations are 2 to 3 times the QL and the high concentrations are 50 times the low concentration. One blank and ten replicates of each matrix at low and high concentrations should be analyzed. The average of the percent of recoveries from ten replicate analyses can be calculated.

Equation 3 $\%R = (C - X)/ T \times 100$

where:
- C = the measured value of the spiked matrix or the standard reference material (SRM)
- X = the indigenous matrix value; zero for the blank matrix
- T = analyte added to the matrix or the certified concentration of the reference material

5.1.5.2 Precision

Precision is a measure of agreement among repeated measurements and provides an estimate of random error. Precision may be assessed by the relative percent difference (RPD) for duplicates and the standard deviation (s), percent relative standard deviation (%RSD, Coefficient of Variation), or the sample range for multiple sample analyses. The data generated from ten replicate analyses of each matrix can be calculated for the method precision.

Equation 4 $RPD = (X_1 - X_2) / \overline{X}$

Equation 5 $RSD = (s / \overline{X}) \times 100$

where:
- s = Standard deviation for ten replicates
- $\overline{X}$ = Mean for 2 or more replicate determinations
- X_1; X_2 = Values of two replicate analyses

For the purpose of technology verification, field replicates should be prepared from a sample taken from one sampling point (split sample rather than co-located sample) for analysis. If the field sample does not contain target analytes, analyses of matrix spike (MS) and matrix spike duplicates (MSDS) are required to provide the precision data. A format for presenting the results of method bias and precision is presented in Table 3.

Table 3 Determination of Method Accuracy and Precision

Sample types	Concentration[1]	*Number of samples*	*Accuracy % Recovery*	*Precision RSD*
Matrix 1	low	10		
	high	10		
Matrix 2	low	10		
	high	10		
Matrix 3	low	10		
	high	10		

[1]The suggested low concentrations are 2 to 3 times the QL, and high concentrations are 50 times the low concentration.

5.1.6 Comparison of Inter-Operator and/or Inter-Laboratory Performance

For a proposed laboratory quantitative method, the technology is required to be operated successfully by at least three independent parties to demonstrate the ruggedness of the method. Six replicate analyses of a performance evaluation sample should be performed by an operator from each laboratory. The mean, the standard deviation, and the coefficient of variation (CV) of six analyses by each operator are calculated to compare the intra-operator and inter-laboratory variabilities. Depending on the analytes and technology, the CV of intra-operator performance should generally not be greater than 30-40%, and the variation of interlaboratory results should not be greater than a factor of 2.

5.1.7 Manufacturer's Publications

Publications related to the proposed technology should be submitted for evaluation. These include product inserts and specifications, user's operating and maintenance manual, application notes, guidance for troubleshooting, and publications in scientific journals, as available and appropriate.

5.2 Field Validation Study

Performance determined by analyzing real-world, contaminated samples is a key element in technology verification. The proponent should identify an independent third party who is responsible for conducting the field verification. The proponent is expected to provide all the required apparatus, equipments and/or materials, as well as operator training. Depending on ability and subject to negotiations, the test plan for the study needs to be prepared either by the proponent or an independent verification entity acceptable to DTSC. Prior to field demonstrations, field testing and quality assurance plans are subject to review by the Department. For technologies requiring advanced instrumentation and/or specialized operating skills, field demonstrations can be conducted by the proponent or the selected party, provided that the demonstration is overseen by DTSC staff or an independent third party acceptable to DTSC. In any case, both the reference laboratory and the field testing are subject to audit while the study is underway. Field demonstration data obtained from participation in Consortium for Site Characterization Technology (CSCT) sponsored by the U.S. EPA are acceptable for evaluation under this program. A guidance manual for the preparation of demonstration plans is available from the U.S. EPA program (Ref. 8).

5.2.1 Site Selection

Sites selected for demonstration should be known to contain target analytes from not-detected to high concentrations. Site information on the distribution of contaminants and the concentration range should be obtained. If this background information is not available, the pre-demonstration testing designed by the proponent or the verification entity must be carried out using a qualified laboratory to determine chemical constituents, the range of concentrations, and the distribution of contaminants at a site. Site control samples which are free from contamination, if available, should be analyzed to check for matrix effects.

5.2.2 Selection of Reference Laboratory and Reference Method

A confirmatory laboratory should be certified by a competent jurisdiction or provide evidence of successful participation in one or several relevant interlaboratory proficiency testing programs. A quality assurance plan and data package for validation must be made available for review by the Department.

Reference methods used for confirmation should have been approved by the U.S. EPA, the Department, or another qualified organization recognized by Cal/EPA. If a regulatory method is not appropriate or is not available, a generally recognized reference method that is responsive to the technology's principle and purpose may be approved by the Department.

5.2.3 Study Design

Based on technology-specific data quality objectives, a field testing plan must be established before environmental sample collection begins. The field testing plan should describe the data quality objectives, the sampling plan, analytical procedures, and strategies used to compare the performance of a test method with the reference method. The decision on error limits and the acceptable data criteria at 90% or 95% confidence levels should be established. The elements of the study design include the intra-method precision and accuracy, the limits for the rates of false positive and false negative results, the representativeness and completeness of the data, the statistics to compare the correlation between the test method and the reference method, and corrective actions needed to be taken under unexpected situations.

5.2.3.1 Sampling Plan

The proponent or the verification entity should assemble a qualified sampling team. Sampling locations which contain low, medium, and high concentrations of target analytes in the matrix of concern should be identified. Sampling techniques and sampling devices must be specified. Adequate amount of environmental sample should be collected and well mixed to prepare split samples for the analyses by both the test method and the reference method. The location and the procedure for preparing the split samples must be documented. Methods of sample preservation, if necessary, and transportation should also be recorded. On the bases of data quality objectives including the tolerance limits of false-positive and false negative results, the confidence level needed, and the budget situation, the minimum number of samples for analyses must be determined (Ref. 9). Ten to twenty percent of field samples should be collected in quadruplicate, with two splits submitted blind for the test method and two splits submitted blind for the reference method. From an economical point of view, it is recommended that an additional 10% of samples be collected in preparation for analyses in case of an unexpected situation.

5.2.3.2 Sample Analysis

For the purpose of method verification, a minimum of thirty split samples which include the concentration range from non-detect to high concentrations should be analyzed by both the test method and the reference method. Field replicates at 10 to 20% of the total samples are required to analyze to access intra-method and inter-method precision. When field samples contain low levels of analyte(s), analyses of duplicates do not provide adequate data on precision. Field duplicates spiked with target analytes at proper concentrations should be prepared. Matrix Spikes (MS) and Matrix Spike Duplicates (MSD) should be analyzed to provide precision data.

For a fully quantitative method, results should be obtained form a five-point calibration curve. For a semiquantitative method measuring multi-analytes, such as BTEX, PAHs, dioxins, PCBs, etc., where the response to the component analytes may vary considerably, it is recommended that a site sample be analyzed and diluted to a proper concentration, usually at a regulatory level, to use as a target level for determining the rate of false positive and false negative results. If the contamination is primarily derived from a single class of known compounds, such as a single Aroclor, etc., this material may be used as a calibrator (Ref. 10).

5.3 Quality Control and Quality Assurance (QC/QA) of the Analysis

Quality assurance plans for the test method by itself and for the field validation study need to be established. The quality assurance project plans required by the U.S. EPA (Ref. 11, 12) or based on the ANSI/ASQC E4 standard (Ref. 13) are recommended for the field demonstration. Such a QC procedure shall include the following elements:

- Instrument calibration criteria must be set. For a fully quantitative test method, the response of a standard at the mid-point of the calibration curve should be recorded to validate the analytical system. For a semiquantitative technique, the acceptance criterion for the response of a calibrator needs to be established by the technology proponent.
- Quality control samples should be analyzed with each batch of environmental samples. These include field blanks, trip blanks (if necessary), method blanks, and check samples.
- If false-negative results are of concern, such as in site investigations, a matrix sample spiked at two times the action level should be included. If false positive results are of concern, such as in remediation, a matrix spiked at one-half the action level should be included.
- A quality control chart should be kept over time to demonstrate the stability of the analytical system.
- Duplicate samples are analyzed with each batch of samples or matrix or one out of every ten samples whichever is more frequent.
- Check samples can be certified reference materials or quality control samples which were prepared from an uncontaminated matrix spiked with known amounts of analytes from a source independent of the calibration standards. The level of the spike shall be at the regulatory action level or the level of concern.
- Chain-of-custody procedures should be in effect throughout the verification study.

Table 4 An Example of Summary of Field Sampling and Analyses

Sample Type	Number of Field Samples		Total Analyses
	Reference Method	Test method	
Field Samples	30	30	60
Field QC Samples			
Field Blank[1]	1	1	2
Field Duplicates[2]	5	5	10
Matrix Spikes	5	5	10
Matrix Spike Duplicates[3]	5	5	10
Check Samples	1	1	2
Travel Blank[4]	1	1	2
Spare Samples[5]	5	5	10
Total	**53**	**53**	**106**

[1]Aliquots of analyte-free sample taken to the field and transferred from the sample container to the individual sample containers in the field as a check on contamination from the atmosphere at the site.
[2]Field duplicates should be collected at 10 to 20% of the samples, unless specified in the quality assurance project plan.
[3]Not analyzed unless the samples contain non-detected or low level analytes.
[4]Trip blank will be prepared but not required to be analyzed unless field blanks indicate a contamination problem.
[5]Extra samples for analyses for unexpected situations.

Quality control procedures for the reference method should follow the instructions given in the U.S. EPA Test Methods for Evaluating Solid Waste (Ref. 14).

6 Data Analysis and Acceptance Criteria

On the bases of the *in-house* performance data and the field testing data, the technology is evaluated against those claims made by the proponent and the proposed or intended use of the technology. The performance standards discussed in the following three sections are applied commonly to evaluate measurement and monitoring technologies unless specified under the performance claims.

6.1 Intramethod Performance

For quantitative methods accuracy, as determined by percent recovery, should be within 70 - 130% of the values spiked into the field matrices. Recoveries less than 70% may be supported by an explanation of method bias. Precision, as measured by relative percent difference (RPD) with a homogeneous matrix, should be less than 40%.

Semiquantitative methods, the numbers of false positive and false negative results relative to a specified action level should be summarized as shown in Table 5. The overall accuracy of the method is calculated as the percentage of samples which are identified correctly. The rates of false positives and false negatives are the percentage of samples which are incorrectly identified as false positive and false negative in the analyses. The overall accuracy, and the rates of false positive and false negative results of the field testing should be checked against the proposed performance standards. 90% of the data are expected to meet the performance claims made by the claimant.

Table 5 The False-Positive and the False-Negative Results

		Reference Method		
		<Target Level	>Target Level	Total
Test Method	<Target Level	a	b (FN)	a + b
	>Target Level	c (FP)	d	c + d
	Total	a + c	b + d	n

Overall accuracy = the percentage of samples which are identified correctly, (a + d) / n
% FP = the percentage of incorrectly identified false positive samples, (c / n) x 100
% FN = the percentage of incorrectly identified false negative samples, (b / n) x 100

6.2 Inter-Method Performance

One objective of verification is to determine whether the data set from the test method can be statistically correlated to that obtained by the reference method, and to compare the accuracy and precision of the two methods. The data generated by the reference method must be verified through critical review according to the QC/QA procedures specified in the U.S. EPA Test Methods for Evaluating Solid Waste (Ref. 13). If results obtained from the reference method were validated, these data would be used as standards to evaluate the performance of the test method. The correlation of field data generated from these two methods is evaluated by statistical methods.

6.2.1 Quantitative Methods

Statistical parameters used to compare data sets include measures of central tendency (mean, median, and mode); measures of dispersion (range, variance, standard deviation, coefficient of variation); and measures of distribution (symmetry or shape). Graphical plots are often used to express the correlation of two sets of data or to monitor the stability of the analytical system over time.

6.2.1.1 Scatter Plot

A scatter plot is generally applied to demonstrate the relationship between paired data points. A scatter plot displays the correlation between two data sets. Potential outliers can be identified by plotting the field data against the values obtained by the reference method.

- Linear Regression

  Linear regression analysis using the method of least squares is most commonly used to demonstrate correlation between the test method and the reference method. For technologies that are capable of producing quantitative results, at least 30 data points are needed for this comparison. Figure 1 (courtesy of Turner Designs) is an example to illustrate the correlation of data between an experimental fluorescence method for petroleum aromatic hydrocarbons in water and the Freon-IR reference method. Linear regression provides the values of y-intercept, slope, and correlation coefficient. A slope of one, a y-intercept of zero, and a correlation coefficient (r^2) of one would mean that the experimental results perfectly match those of the reference method. Because of the heterogeneous nature of environmental samples, r^2 values greater than 0.85 have been considered meeting data quality requirements for laboratory analysis, r^2 values greater than 0.75 have been considered meeting data quality requirements for field analysis.

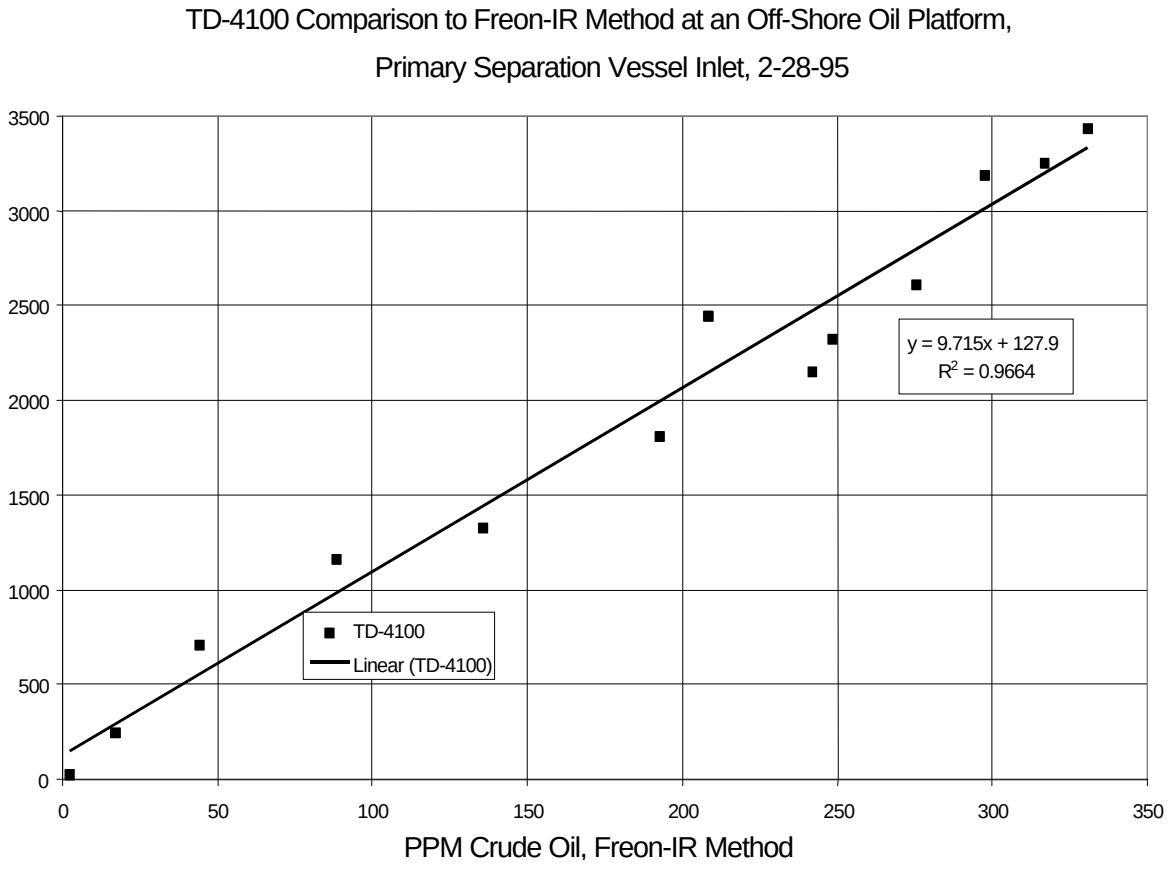


Figure 1 Correlation of Two Sets of Data by Linear Regression Analysis

- Method Bias

  The existence of a bias between the test method and the reference method may be illustrated by plotting the relative percent difference (RPD) between values generated by the test method and the reference method versus the concentration determined by the reference method as illustrated in Figure 2.

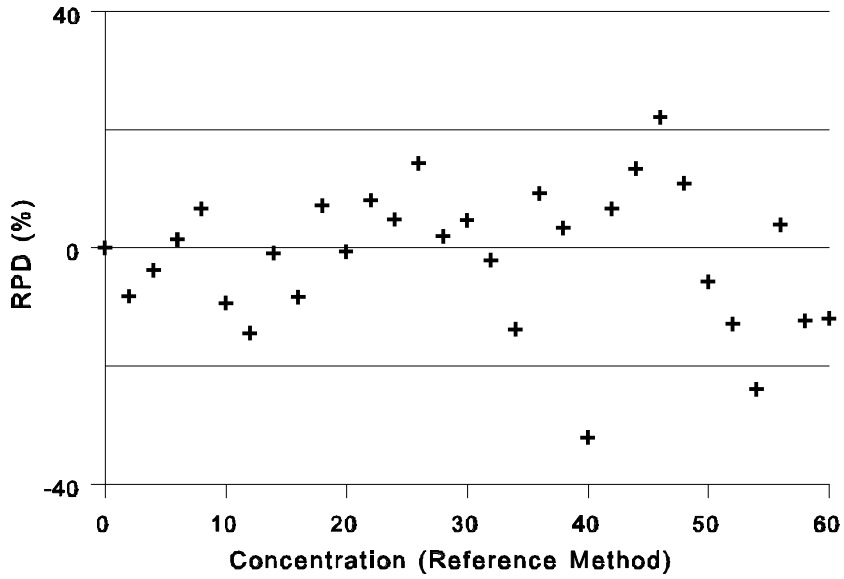


Figure 2 Method Bias vs. Concentration

On the basis that the result obtained from the reference method is accurate, RPD is calculated as ((y_i - x_i) / x_i) x (100). The horizontal center line of this plot has the value of zero, indicating no difference between the results from the test method and the reference method. The bias between these two methods appears on the vertical axis as relative percent difference. This figure would illustrate the positive or negative bias of a test method with respect to the reference method as a function of concentration.

6.2.1.2 Statistical Analysis

The correlation of field data obtained from the test method and from the reference method should be assessed by statistical methods. Discussions on the selection of appropriate statistical hypothesis tests to evaluate the correlation of the two data sets can be found in the EPA publication Guidance for Data Quality Assessment (Ref. 14). For simple random sampling and randomized systematic sampling designs, provided that two sample means are approximately normally distributed, the two-sample Student t-test and the F-test are commonly used to check for the similarity of two data sets. Descriptions for the Student t-tests to compare two population means and the F-test to compare two sample variances are attached in Appendix B and Appendix C for reference.

6.2.2 Semiquantitative Methods

The comparison of performance data on semi-quantitative methods would be based on the design of the analytical scheme. Table 6 is an example of presenting a semi-quantitative method with respect to the numbers of samples tested by the semi-quantitative method and the reference method.

Table 6 Comparison of Semi-quantitative Data and Confirmatory Data

		EPA Method 8270A (ppm)				
Test Method (ppm)		<1	>1<10	>10<100	>100	Total
	<1	12	7	0	0	9
	>1<10	2	11	3	0	16
	>10<100	0	10	7	0	17
	>100	0	1	8	53	62
	Total	14	29	18	53	114

The column and row headings reflect that, in semi-quantitative methods, sample concentrations are typically classified as ranges rather than fully quantitative values. Cells found along the diagonal are samples identified correctly by both the test method and the reference method. Assuming that all results obtained by the reference method are accurate, samples located above the diagonal are underestimated in terms of concentration, and those below the diagonal are overestimated.

Figure 3 illustrates the distribution of false-positive and false-negative results obtained by a screening method for a certain type of soil contaminant at a target level of 100 ppm.

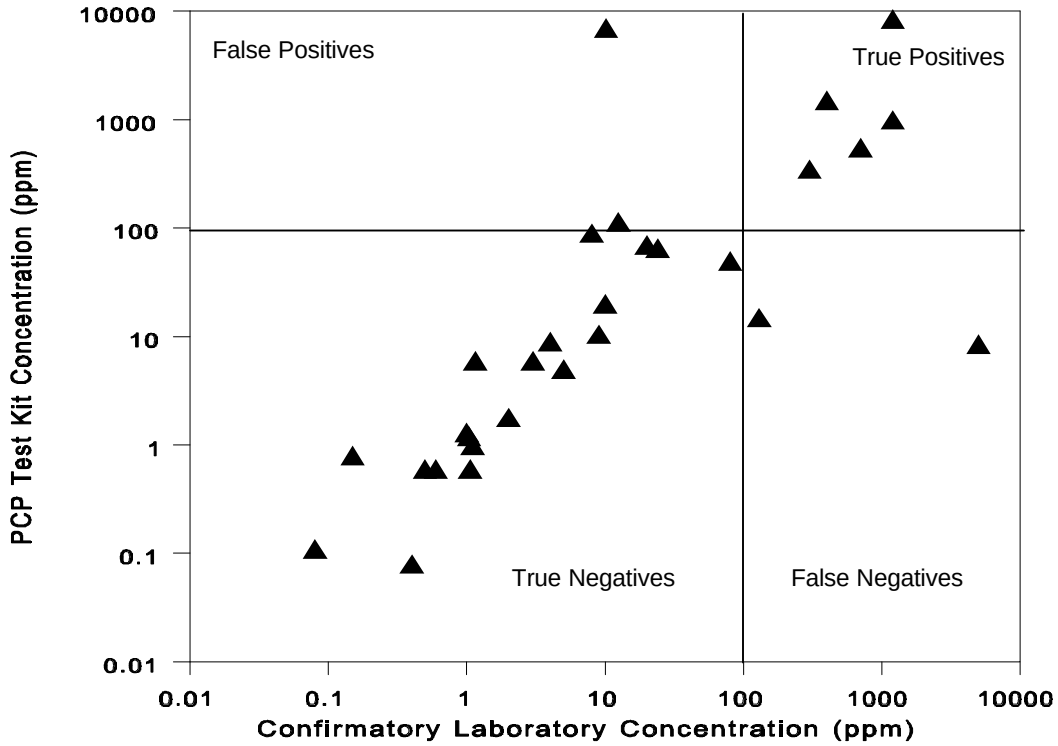


Figure 3 Comparison of Performance Data Between the Test Method and the Reference Method with Respect to the Regulatory Action Level

6.3 Assessment of Existing Data

Existing field data can be accepted for technology verification provided that the field testing was conducted under the same requirements as stated in section 5.2 *Field Validation Study* and data quality meets the following criteria:

- The field testing was conducted by an independent third party acceptable to the Department or by the proponent under the supervision of DTSC staff. The name of the organization, and the contact person who was responsible for the study are available for information as needed. The history of site usage, contamination profile, and the name of other organizations who were involved in this study are documented.
- Documentation includes the dates of field testing, the purpose of the study, and the study design which describes the sampling procedures, sample analysis, and quality control and quality assurance procedures used for the field and the laboratory testing.
- The data will be evaluated only if they were generated according to the conditions specified in the study design, and the QC data generated under the same analytical conditions fell within the range of acceptance.
- The confirmatory laboratory used for the method validation should meet the qualifications stated in 5.2.2.
- The submitted data were relevant to the performance claims set forth by the proponent.

According to the described criteria the existing data can be accepted partially or completely for the evaluation. The proponent is required to generate the additional data as needed for the technology verification.

7 Applications

On the basis of the results obtained, a technology can be characterized as to its capability and limitations relative to each intended application. This includes its applicability on analytes and their concentrations; media types and matrices; environmental and operating conditions. If the technology can be reasonably expected to perform under conditions not encountered in the testing, such application can be stated with a requirement for confirmatory testing before actual use. As a reasonable warning to users, conditions and areas of application for which the technology is known to be unsuitable, will also be included.

8 Reliability

The reliability of a technology is its ability to perform acceptably over time. Indicators of reliable performance and indicators of malfunction should be established. Generally, tests of reliability should have been included in the proponent's product development record. Information on holding times of samples, the shelf life of reagents and other materials must be provided. For new technologies, approaches for troubleshooting, and safeguards against performance failure should be provided by the proponent. If available, the company's warranty policy must be presented and a user hot-line and other customer support provided.

9 Quality Control in Manufacturing

A technology proponent needs to disclose to Cal/EPA the quality assurance program that is in effect in the design, manufacture, distribution, and service of the technology. Quality control limits, acceptance criteria, monitoring procedures for manufacture, and maintenance requirements need to be summarized and documentation provided upon request. Internal QA data and data with proprietary information which is properly identified as such will be withheld from the public record to the extent provided by law. Manufacturer's certification under the ISO-9000 series of standards should be indicated in the application. (ISO-9000 certification or maintenance of an ISO-9000 certifiable quality system will facilitate DTSC's task of monitoring the quality of a certified technology as mandated by law.) By accepting certification, the proponent enters a commitment to maintain the technology's quality standards over the period of the certification.

10 Safety

The evaluation of health and safety issues is mandated by law. A technology proponent needs to identify the health hazard or risk that could potentially impact the operator, the public, and the environment. These include the safety precautions related to equipment operation, exposures to toxic chemicals, etc. The Material Safety Data Sheet (MSDS) and instructions for health and safety must be provided. The procedure for waste disposal must be stated.

11 Training Requirements

The operator's training requirements and qualifications must be stated. In consideration of the successful operation of the technology and safety issues, the company should provide "on-site" and/or "off-site" training as needed.

EVALUATION REPORT

A Technology Evaluation Report prepared by Department staff will summarize the results of this evaluation and name the recommended applications for the technology, and the technology's limitations. For "screening" methods, a certain frequency of confirmation by a fully qualitative and quantitative reference method will be recommended, except that these may be modified by the data quality objectives specified in the quality assurance project plan. A Certification Statement summarizing the operating principles, performance evaluations, and the suggested applications of the technology would be published in the California Regulatory Notice Register.

CHECK LIST FOR DATA PACKAGE AND INFORMATION

Data Package and Information	***Complete***	***Incomplete***
A. Technology Description		
Source of scientific and engineering principles		
Performance claim(s)		
Scope of applications		
Basis for performance criteria (standards set for the reference method; levels of regulatory limit or risk-based concern; or specified project requirements)		
Reference method selected for the comparison of performance, (reference source, method number, title, and date, if applicable)		
B. Performance Data		
Standard operating procedures (sampling, sample preparation, analytical procedures, --- , as appropriate)		
Proposed target analyses		
Concentration of calibration standards of each analyte		
Slope, intercept, and correlation of calibration regression (quantitative method only)		
Source of standard, and spiking material used		
Media and matrices used for the study		
Performance range tested (with units)		
Method detection limit (spiked sample concentration; number of replicate analyses of each analyte)		
Quantitation limit (quantitative method only)		
Precision (spiked sample concentrations at high and low levels of each analyte; number of replicate analyses)		
Bias (spiked sample concentrations at high and low concentrations of each analyte; number of replicate analyses, quantitative method only)		
Surrogates used, if applicable (spiked concentration and recovery data, quantitative method only)		
Rates of the false positive and the false negative results (screen method only)		

Specificity (if applicable)		
Interferences (and cross reactivities, if applicable)		
Training requirements (if necessary)		
Performance data of reference method		
Accreditation of the confirmatory laboratory		
C. QA/QC Data		
Qualitative identification criteria		
Instrument calibration criteria		
Result(s) of method blank		
Results of laboratory QC samples and/or PE samples with certified value and range of acceptance		
Duplicate analyses (one of every ten samples; for samples less than ten one of each batch or matrix)		
Chain of custody records		
D. Data Reporting		
Quantitative method (MDL, QL, method bias and precision, statistical data to demonstrate the correlation of the test method and the reference method)		
Screening method (data of lowest detection level, method precision, rates of false positive and false negative results relative to the action level or the project specified target level)		
E. Existing Performance Data (if available)		
Validation data by third parties (sponsor organization, project title, location and date of the study)		
Documentation of study design and QC data of field and laboratory analyses		
F. Summary of Technology Performance		

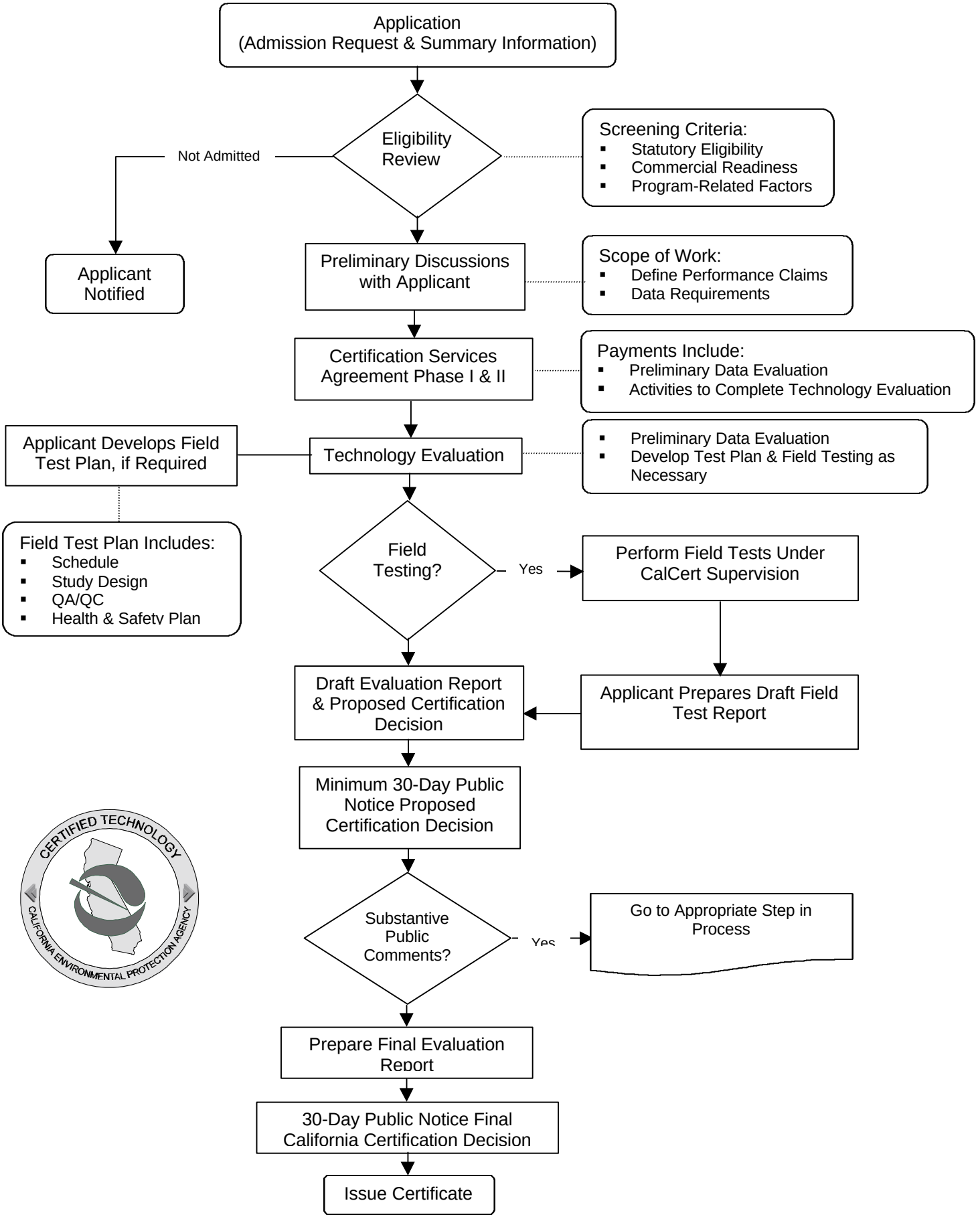
REFERENCES

1. Guidance for Method Development and Methods Validation for the RCRA Program, Methods Section (5304), April 6, 1992, U.S. EPA, Office of Solid Waste, Washington, DC 20460.
2. Test Methods for Evaluating Solid Waste, Volume 1A: Laboratory Physical/Chemical Methods, Chapter one quality control, Method Detection Limit, p. 25-27, Revision 1, July 1992, U.S. EPA, Office of Solid Waste and Emergency Response, Washington, DC 20460.
3. Title 40, Code of Federal Regulations, Protection of Environment (40 CFR, part 261) July 1, 1992.
4. Title 22, CCR, Chapter11, Division 4.5 Environmental Health Standards for the Management of Hazardous Waste (Div 4.5. 22 CCR, Sec 66261) April 3, 1998.
5. Preliminary Remediation Goals (PRGs), U.S. EPA Region 9, August, 1996.
6. Incident of Adverse Biological Effects Within Ranges of Chemical Concentrations in Marine Sediments, by E. R. Long et al. Environmental Management Vol. 19, No.1, p. 81-97.
7. U.S. EPA Test Methods for Evaluating Solid Waste, Volume 1A: Laboratory Physical/Chemical Methods, Update III, Method 4000, Revision 0, December 1996, U.S. EPA, Office of Solid Waste and Emergency Response, Washington, DC 20460.
8. A Guidance Manual for the Preparation of Site Characterization and Monitoring Technology Demonstration Plans, Interim Final October, 1996, U.S. EPA, Office of Research and Development, National Exposure research Laboratory, Characterization Research Division, Las Vegas, NV.
9. A Guidance for Data Quality Objectives Process, EPA QA/G-4, EPA/600/R-96/055, September 1994, U.S. EPA, Office of Research and Development, Washington, DC 20460
10. The RaPID Assay Environmental User's Guide (Rev. 12/94), Strategic Diagnostics, Inc. Newark, DE.
11. EPA Requirements for Quality Assurance Project Plans for Environmental Data Operators, EPA QA/R-5, U.S. EPA, Quality Assurance Division, Washington, DC 20460, Draft-November 1997.
12. EPA Guidance for Quality Assurance Project Plans, EPA QA/G-5, EPA/600/R-98/018, February 1998.
13. American Society for Quality Control, "Specification and Guidance for Quality System for Environmental Data Collection and Environmental Technology Programs," American National Standards, ANSI/ASQC E4 1994.
14. U.S. EPA Test Methods for Evaluating Solid Waste, Update 3, Chapter One, 3.5 Quality Assurance Review, Revision 1, July 1, 1992, U.S. EPA, Office of Solid Waste and Emergency Response, Washington, DC 20460.
15. Guidance for Data Quality Assessment, Practical Methods for Data Analysis, EPA QA/G-9, EPA/600/R-96/084, July 1966, U.S. EPA, Office of Research and Development, Washington, DC 20460.

ACKNOWLEDGMENTS

This draft document incorporated the comments and suggestions made by our department members, Dr. Barton P. Simmons and Cindy M. Dingman of the Department of Toxic Substances Control, Hazardous Materials Laboratory; Dr. Stephen Billets and Eric N. Koglin of U.S. Environmental Protection Agency, National Exposure Research Laboratory in Las Vegas, NV; Derek Vachon of Water Technology International Corporation, Burlington, Ontario, Canada; and stakeholders participating in workshops on Environmental Technology Verification. We appreciate their contributions to this test protocol.

Hazardous Waste Environmental Technologies – A Typical Certification Process



APPENDIX A Method Detection Limit (Ref. 2)

The method detection limit (MDL) is the minimum concentration of a substance that can be measured and reported with 99% confidence that the analyte concentration is greater than zero and is determined from analysis of a sample in a given matrix type containing the analyte.

For operational purposes, when it is necessary to determine the MDL in the matrix, the MDL should be determined by multiplying the appropriate one-sided 99% t-statistic by the standard deviation obtained from a minimum of three analyses of a matrix spike containing the analyte of interest at a concentration three to five times the estimated MDL, where the t-statistic is obtained from standard references or the table below.

No of samples	t-statistic
3	6.96
4	4.54
5	3.75
6	3.35
7	3.14
8	3.00
9	2.90
10	2.82

Estimate the MDL as follows:

Obtain the concentration value that corresponds to:

a) an instrument signal/noise ratio within the range of 2.5 to 5.0, or

b) the region of the standard curve where there is a significant change in sensitivity

Determine the variance (S^2) for each analyte as follows:

$$S^2 = [1/(n-1)][\sum_{i=1}^{n}(x_i-\bar{x})^2]$$

where x_i = the ith measurement of the variable x

and $\bar{x}$ = the average value of x

Determine the standard deviation (s) for each analyte as follows:

$$s = (S^2)^{1/2}$$

Determine the MDL for each analyte as follows:

$$MDL = t_{(n-1,\ a = .99)}(S)$$

where $t_{(n-1,\ a = .99)}$ is the one-sided t-statistic appropriate for the number of samples used to determine the standard deviation (s), at the 99 percent level.

APPENDIX B The F-test for the Equality of two Variances (Ref. 15, 4.5.2)

An F- test may be used to test whether the true underlying variances of two population are equal. Usually the F-test is employed as a preliminary test, before conducting the two-samples t-test for the equality of two means. The assumptions underlying the F-test are that two samples are independent random samples for two underlying normal populations. Directions for implementing an F-test with an example are given in Box 4.5-2.

Box 4.5-2: Directions for Calculating an F-Test to Compare Two Variances with an Example

Directions: Let $X_1, X_2, \ldots, X_m$ represent the m data points from population 1 and $Y_1, Y_2, \ldots, Y_n$ represent the n data points from population 2. To perform an F-test, proceed as follows.

STEP 1: Calculate the sample variances s_{X^2} (for the X's) and s_{Y^2} (for the Y's) (section 2.2.3).

STEP 2: Calculate the variance ratios $F_X = s_{X^2}/s_{Y^2}$ and $F_Y = s_{Y^2}/s_{X^2}$. Let F equal the larger of these two values. If $F = F_x$, then let k = m - 1 and q = n - 1. If $F = F_y$, then let k = n -1 and q = m - 1.

STEP 3: Using Table A-9 of Appendix A of the F distribution, find the cutoff $U = f_{1-"/2}(k, q)$. If F > U, conclude that the variances of the two populations are not the same.

Example: Manganese concentrations were collected from 2 wells. The data are Well X: 50, 73, 244, and 202 ppm; and Well Y: 272, 171, 32, 250, and 53 ppm. An F-test will be used to determine if the variances of the two wells are equal.

STEP 1: For Well X, $s_{X^2} = 9076$. For Well Y, $s_{Y^2} = 12125$.

STEP 2: $F_X = s_{X^2}/s_{Y^2} = 9076 / 12125 = 0.749$. $F_Y = s_{Y^2}/s_{X^2} = 12125 / 9076 = 1.336$. Since, $F_Y > F_X$, $F = F_Y = 1.336$, k = 5 -1 = 4 and q = 4 - 1 = 3.

STEP 3: Using Table A-9 of Appendix A of the F distribution with " = 0.05, $U = f_{1-.05/2}(4, 3) = 15.1$. Since 1.336 < 15.1, there is no evidence that the variability of the two wells is different.

(This page intentionally left blank.)

Table A-9: Percentiles of the F Distribution

Degrees Freedom for Denominator	Degrees of Freedom for Numerator																	
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	60	120	∞
1																		
.50	1.00	1.50	1.71	1.82	1.89	1.94	1.98	2.00	2.03	2.04	2.07	2.09	2.12	2.13	2.15	2.17	2.18	2.20
.90	39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2	60.7	61.2	61.7	62.0	62.3	62.8	63.1	63.3
.95	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	252	253	254
.975	648	800	864	900	922	937	948	957	963	969	977	985	993	997	1001	1010	1014	1018
.99	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6235	6261	6313	6369	6366
2																		
.50	0.667	1.00	1.13	1.21	1.25	1.28	1.30	1.32	1.33	1.34	1.36	1.38	1.39	1.40	1.41	1.43	1.43	1.44
.90	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.48	9.49
.95	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5
.975	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5	39.5
.99	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5
3																		
.50	0.585	0.881	1.00	1.06	1.10	1.13	1.15	1.16	1.17	1.18	1.20	1.21	1.23	1.23	1.24	1.25	1.26	1.27
.90	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.15	5.14	5.13
.95	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.87	8.79	8.74	8.70	8.66	8.64	8.62	8.57	8.55	8.53
.975	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.3	14.3	14.2	14.1	14.1	14.0	13.9	19.3
.99	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.3	26.2	26.1
4																		
.50	0.549	0.828	0.941	1.00	1.04	1.06	1.08	1.09	1.10	1.11	1.13	1.14	1.15	1.16	1.16	1.18	1.18	1.19
.90	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.79	3.78	3.76
.95	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.69	5.66	5.63
.975	12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.36	8.31	8.26
.99	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.6	13.5
.999	74.1	61.2	56.2	53.4	51.7	50.5	49.7	49.0	48.5	48.1	47.4	46.8	46.1	45.8	45.4	44.7	44.4	44.1
5																		
.50	0.528	0.799	0.907	0.965	1.00	1.02	1.04	1.05	1.06	1.07	1.09	1.10	1.11	1.12	1.12	1.14	1.14	1.15
.90	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.39	3.27	3.24	3.21	3.19	3.17	3.14	3.12	3.11
.95	6.61	5.79	5.41	5.19	5.05	1.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.43	4.40	4.37
.975	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.73	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.12	6.07	6.02
.99	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.20	9.11	9.02
.999	47.2	37.1	33.2	31.1	29.8	28.8	28.2	27.6	27.2	26.9	26.4	25.9	25.4	25.1	24.9	24.3	24.1	23.8
6																		
.50	0.515	0.780	0.886	0.942	0.977	1.00	1.02	1.03	1.04	1.05	1.06	1.07	1.08	1.09	1.10	1.11	1.12	1.12
.90	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.76	2.74	2.72
.95	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.74	3.70	3.67
.975	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	4.96	4.90	4.85
.99	22.8	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.06	6.97	6.88
.999	35.5	27.0	23.7	21.9	20.8	20.0	19.5	19.0	18.7	18.4	18.0	17.6	17.1	16.9	16.7	16.2	16.0	15.7

EPA QA/G-9 QA96

Table A-9: Percentiles of the F Distribution

Degrees Freedom for Denominator	Degrees of Freedom for Numerator																	
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	60	120	∞
7																		
.50	0.506	0.767	0.871	0.926	0.960	0.983	1.00	1.01	1.02	1.03	1.04	1.05	1.07	1.07	1.08	1.09	1.10	1.10
.90	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.51	2.49	2.47
.95	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.30	3.27	3.23
.975	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.25	4.20	4.14
.99	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.82	5.74	5.65
.999	29.2	21.7	18.8	17.2	16.2	15.5	15.0	14.6	14.5	14.1	13.7	13.3	12.9	12.7	12.5	12.1	11.9	11.7
8																		
.50	0.499	0.757	0.860	0.915	0.948	0.971	0.988	1.00	1.01	1.02	1.03	1.04	1.05	1.06	1.07	1.08	1.08	1.09
.90	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.34	2.32	2.29
.95	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.01	2.97	2.93
.975	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.78	3.73	3.67
.99	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.03	4.95	4.86
.999	25.4	18.5	15.8	14.4	13.5	12.9	12.4	12.0	11.8	11.5	11.2	10.8	10.5	10.3	10.1	9.73	9.53	9.33
9																		
.50	0.494	0.749	0.852	0.906	0.939	0.962	0.978	0.990	1.00	1.01	1.01	1.03	1.04	1.05	1.05	1.07	1.07	1.08
.90	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.21	2.18	2.16
.95	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.79	2.75	2.71
.975	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.45	3.39	3.33
.99	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.48	4.40	4.31
.999	22.9	16.4	13.9	12.6	11.7	11.1	10.7	10.4	10.1	9.89	9.57	9.24	8.90	8.72	8.55	8.19	8.00	7.81
10																		
.50	0.490	0.743	0.845	0.899	0.932	0.954	0.971	0.983	0.992	1.00	1.01	1.02	1.03	1.04	1.05	1.06	1.06	1.07
.90	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.11	2.08	2.06
.95	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.84	2.77	2.74	2.70	2.62	2.58	2.54
.975	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.20	3.14	3.08
.99	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.08	4.00	3.91
.999	21.0	14.9	12.6	11.3	10.5	9.93	9.52	9.20	8.96	8.75	8.45	8.13	7.80	7.64	7.47	7.12	6.94	6.76
12																		
.50	0..484	0.735	0.835	0.888	0.921	0.943	0.959	0.972	0.981	0.989	1.00	1.01	1.02	1.03	1.03	1.05	1.05	1.06
.90	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.96	1.93	1.90
.95	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.38	2.34	2.30
.975	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.85	2.79	2.72
.99	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.54	3.45	3.36
.999	18.6	13.0	10.8	9.63	8.89	8.38	8.00	7.71	7.48	7.29	7.00	6.71	6.40	6.25	6.09	5.76	5.59	5.42
15																		
.50	0.478	0.726	0.826	0.878	0.911	0.933	0.949	0.960	0.970	0.977	0.989	1.00	1.01	1.02	1.02	1.03	1.04	1.05
.90	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.82	1.79	1.76
.95	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.16	2.11	2.07
.975	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.52	2.46	2.40
.99	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.05	2.96	2.87
.999	16.6	11.3	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	5.81	5.54	5.25	5.10	4.95	4.64	4.48	4.31

Table A-9: Percentiles of the F Distribution

Degrees Freedom for Denominator	Degrees of Freedom for Numerator																	
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	60	120	∞
20																		
.50	0.472	0.718	0.816	0.868	0.900	0.922	0.938	0.950	0.959	0.966	0.977	0.989	1.00	1.01	1.01	1.02	1.03	1.03
.90	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.68	1.64	1.61
.95	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.95	1.90	1.84
.975	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.22	2.16	2.09
.99	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	30.9	2.94	2.86	2.78	2.61	2.52	2.42
.999	14.8	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.82	4.56	4.29	4.15	4.00	3.70	3.54	3.38
24																		
.50	0.469	0.714	0.812	0.863	0.895	0.917	0.932	0.944	0.953	0.961	0.972	0.983	0.994	1.00	1.01	1.02	1.02	1.03
.90	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.61	1.57	1.53
.95	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.84	1.79	1.73
.975	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.08	2.01	1.94
.99	7.82	6.66	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.40	2.31	2.21
.999	14.0	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80	4.64	4.39	4.14	3.87	3.74	3.59	3.29	3.14	2.97
30																		
.50	0.466	0.709	0.807	0.858	0.890	0.912	0.927	0.939	0.948	0.955	0.966	0.978	0.989	0.944	1.00	1.01	1.02	1.02
.90	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.62	1.64	1.61	1.54	1.50	1.46
.95	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.74	1.68	1.62
.975	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	1.94	1.87	1.79
.99	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.21	2.11	2.01
.999	13.3	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	4.00	3.75	3.49	3.36	3.22	2.92	2.76	2.59
60																		
.50	0.461	0.701	0.798	0.849	0.880	0.901	0.917	0.928	0.937	0.945	0.956	0.967	0.978	0.983	0.989	1.00	1.01	1.01
.90	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.40	1.35	1.29
.95	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.53	1.47	1.39
.975	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.67	1.58	1.48
.99	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.30	1.84	1.73	1.60
.999	12.0	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.69	3.54	3.32	3.08	2.83	2.69	2.55	2.25	2.08	1.89
120																		
.90	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.32	1.26	1.19
.95	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.43	1.35	1.25
.975	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.95	1.82	1.76	1.69	1.53	1.43	1.31
.99	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.66	1.53	1.38
.999	11.4	7.32	5.78	4.95	4.42	4.04	3.77	3.55	3.38	3.24	3.02	2.78	2.53	2.40	2.26	1.95	1.77	1.54
∞																		
.90	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.24	1.17	1.00
.95	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.32	1.22	1.00
.975	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.39	1.27	1.00
.99	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.47	1.32	1.00
.999	10.80	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10	2.96	2.74	2.51	2.27	2.13	1.99	1.66	1.45	1.00

(This page intentionally left blank.)

APPENDIX C Student's Two-Sample t-Test (Ref. 15)

Tests for Comparing Two Populations

A two-sample test involves the comparison of two populations or a "before and after" comparison. In environmental applications, the two populations to be compared may be a potentially contaminated area with a background area or concentration levels from an upgradient and a downgradient well. The comparison of the two populations may be based on a statistical parameter that characterizes the relative location (e.g., a mean or median), or it may be based on a distribution-free comparison of the two population distributions. Tests that do not assume an underlying distributions (e.g., normal or lognormal) are called distribution-free or nonparametric tests. These tests are often more useful for comparing two populations than those that assume a specific distribution because they make less stringent assumptions. Section 3.3.1 covers tests for differences in the means of two populations. Section 3.32 covers tests for differences in the proportion or percentiles of two populations. Section 3.3.3 describes distribution-free comparisons. Section 3.3.4 describes tests for comparing two medians.

Often, a two-sample test involves comparing of the difference of two population parameters to a threshold value. For environmental applications, the threshold value is often zero, representing the case where the data are used to determine which of the two population parameters is greater than the other. For example, concentration levels from a Superfund site may be comapred to a background site. Then, if the Superfund site levels exceed the background levels, the site requires further investigation. A two-sample test may also be used to compare readings from two instruments or two separate populations of people.

If the exact same sampling locations are used for both populations, then the two samples are not independent. This case should be converted to a one-sample problem by applying the methods described in Section 3.2 to the differences between the two populations at the same location. For example, one could compare contaminant levels from several wells after treatment to contaminant levels from the same wells before treatment. The methods described in Section 3.2 would then be applied to the differences between the before and after treatment contaminant levels for each well.

3.3.1 Comparing Two Means

Let μ_1 represent the mean of population 1 and μ_2 represent the mean of population 2. The hypotheses considered in this section are:

Case 1: H_0: $\mu_1 - \mu_2 \leq \delta_0$ vs. H_A: $\mu_1 - \mu_2 > \delta_0$; and
Case 2: H_0: $\mu_1 - \mu_2 \geq \delta_0$ vs. H_A: $\mu_1 - \mu_2 < \delta_0$.

An example of a two-sample test for population means is comparing the mean contaminant level at a remediated Superfund site to a background site; in this case, δ_0 would be zero. Another example is a Record of Decision for a Superfund site which specifies that the remediation technique must reduce the mean contaminant level by 50 ppm each year. Here, each year would be considered a separate population and δ_0 would be 50 ppm.

The information required for these tests includes the null and alternative hypotheses (either Case 1 or Case 2); the gray region (i.e., a value $\delta_1 > \delta_0$ for Case 1 or a value $\delta_1 < \delta_0$ for Case 2 representing the bound of the gray region); the false positive error rate α at δ_0; the false negative error rate β at δ_1; and any additional limits on decision errors. It may be helpful to label additional false positive error limits as α_2 at $\delta_{\alpha 2}$, $\alpha 3$ at $\delta_{\alpha 3}$, etc., and to label additional false negative error limits as β_2 at $\delta_{\beta 2}$, β_3 at $\delta_{\beta 3}$, etc.

3.3.1.1 Student's Two-Sample t-Test (Equal Variances)

Purpose

Student's two-sample t-test can be used to compare two population means based on the independent random samples $X_1, X_2, \ldots, X_m$ from the first population, and $Y_1, Y_2, \ldots, Y_n$ from the second population. This test assumes the variabilities (as expressed by the variance) of the two populations are approximately equal. If the two variances are not equal (a test is described in Section 4.5), use Satterthwaite's t test (Section 3.3.1.2).

Assumptions and Their Verification

The principal assumption required for the two-sample t-test is that a random sample of size m ($X_1, X_2, \ldots, X_m$) is drawn from population 1, and an independent random sample of size n ($Y_1, Y_2, \ldots, Y_n$) is confirmed by reviewing the procedures used to select the sampling points.

The second assumption required for the two-sample t-tests are that the sample means $\overline{x}$ (sample 1) and $\overline{Y}$ (sample 2) are approximately normally distributed. If both m and n are large, one may make this assumption without further verification. For small sample sizes, approximate normality of the sample means can be checked by testing the normality of each of the two samples.

Limitations and Robustness

The two-sample t-test with equal variances is robust to violations of the assumptions of normality and equality of variances. However, if the investigator has tested and rejected normality or equality of variances, then nonparametric procedures may be applied. The t-test is not robust to outliers because sample means and standard deviations are sensitive to outliers.

Sequence of Steps

Directions for the two-sample t-test for a simple random sample and a systematic simple random sample are given in Box 3.3-1 and an example in Box 3.3-2.

3.3.1.2 Satterthwaite's Two-Sample t-test (Unequal Variances)

Satterthwaite's t-test should be used to compare two population means when the variances of the two populations are not equal. It requires the same assumptions as the two-sample t-test (Section 3.3.1.1) except the assumption of equal variances.

Directions for Sattethwaite's t-test for a simple random sample and a systematic simple random sample are given in Box 3.3-3 and an example in Box 3.3-4.

Box 3.3-1: Directions for the Student's Two-Sample t-Test (Equal Variances) for Simple and Systematic Random Samples

This describes the steps for applying the two-sample t-tests for differences between the population means when the two population variances are equal for Case 1 (H_0: $\mu_1 - \mu_2 \leq \delta_0$). Modifications for Case 2 (H_0: $\mu_1 - \mu_2 \geq \delta_0$) are given in parentheses {}.

STEP 1: Calculate the sample mean $\overline{X}$ and the sample variance s_x^2 for sample 1 and compute the sample mean $\overline{Y}$ and the sample variance s_y^2 for sample 2.

STEP 2: Use section 4.5 to determine if the variances of the two populations are equal. If the variances of the two populations are not equal, use Satterthwaite's t test (section 3.3.1.2). Otherwise, compute the pooled standard deviation

$$S_E = \sqrt{\frac{(m-1)s_x^2 + (n-1)s_y^2}{(m-1)+(n-1)}}$$

STEP 3: Calculate

$$t = \frac{\overline{X} - \overline{Y} - \delta_0}{S_E\sqrt{1/n + 1/m}}$$

Use Table A-1 of Appendix A to find the critical value t_{1-a} such that 100(1-a)% of the t-distribution with (m+n-2) degrees of freedom is below t_{1-a}.

If $t > t_{1-a}$ {$t < -t_{1-a}$}, the null hypothesis may be rejected. Go to STEP 5.

If $t \not> t_{1-a}$ {$t \not< -t_{1-a}$}, there is not enough evidence to reject the null hypothesis. Therefore, the false negative error rate will need to be verified. Go to STEP 4.

STEP 4: To calculate the power of the test, assume that the true values for the mean and standard deviation are those obtained in the sample and use a statistical software package like the DEFT software (EPA G-4D, 1994) or the DataQUEST software (EPA G-9D, 1996) to generate the power curve of the two-sample t-test. If only one false negative error rate (β) has been specified (at δ_1), it is possible to calculate the sample size which achieves the DQOs, assuming the true mean and standard deviation are equal to the values estimated from the sample, instead of calculating the power of the test. Calculate

$$m^* = n^* = \frac{2s^2(z_{1-a}+z_{1-\beta})^2}{(\delta_1-\delta_0)^2} + (0.25)z_{1-a}^2$$

If $m^* \leq m$ and $n^* \leq n$, the false negative error rate has been satisfied. Otherwise, the false negative error rate has not been satisfied.

STEP 5: The results of the test could be:

1) the null hypothesis was rejected, and it seems $\mu_1 - \mu_2 > \delta_0$ {$\mu_1 - \mu_2 < \delta_0$};
2) the null hypothesis was not rejected, the false negative error rate was satisfied, and it seems $\mu_1 - \mu_2 \leq \delta_0$ {$\mu_1 - \mu_2 \geq \delta_0$}; or
3) the null hypothesis was not rejected, the false negative error rate was not satisfied, and it seems $\mu_1 - \mu_2 \leq \delta_0$ {$\mu_1 - \mu_2 \geq \delta_0$}, but this conclusion is uncertain because the sample was too small.

Box 3.3-2: An Example of a Student's Two-Sample t-Test (Equal Variances) for Simple and Systematic Random Samples

At a hazardous waste site, area 1 (cleaned using an in-situ methodology) was compared with a similar (but realitvely uncontaminated) reference area, area 2. If the in-situ methodology worked, then the two sites should be approximately equal in average contaminant levels. If the methodology did not work, then area 1 should have a higher average than the reference area. Seven random samples were taken from area 1, and eight were taken from area 2. Because the contaminant concentrations in the two areas are supposedly equal, the null hypothesis is H0: $\mu 1 - \mu 2 \leq 0$ (Case 1). The false positive error rate was set at 5% and the false negative error rate was set at 20% (β) if the difference between the areas is 2.5ppb.

STEP 1:

	Sample Mean	Sample Variance
Area 1	7.8 ppm	2.1 ppm^2
Area 2	6.6 ppm	2.2 ppm^2

STEP 2: Methods described in Section 4.5 were used to determine that the variances were essentially equal. Therefore,

$$S_E = \sqrt{\frac{(7-1)2.1 + (8-1)2.2}{(7-1) + (8-1)}} = 1.4676$$

STEP 3:

$$t = \frac{7.8 - 6.6 - 0}{1.4676\sqrt{1/7+1/8}} = 1.5798$$

Table A-1 of Appendix A was used to find that the critical value $t_{0.95}$ with (7+8-2)=13 degrees of freedom is 1.771.

Because $t \not> t_{1-a}$ (i.e., 1.5798 $\not>$ 1.771), there is not enough evidence to reject the null hypothesis. The false negative error rate will need to be verified.

STEP 4: Assuming the true values for the mean and standard deviation are those obtained in the sample:

$$m^* = n^* = \frac{2(1.4676^2)(1.645 + 0.842)^2}{(2.5 - 0)^2} + (0.25)1.645^2 = 4.938, \text{ i.e., } 5$$

Because $m^* \leq m$ (7) and $n^* \leq n$ (8), the false negative error rate has been satisfied.

STEP 5: The null hypothesis was not rejected and the false negative error rate was satisfied. Therefore, it seems there is no difference between the two areas and that the in-situ methodology worked as expected.

Box 3.3-3: Directions for Satterthwaite's t-Test (Unequal Variances) for Simple and Systematic Random Samples

This describes the steps for applying the two-sample t-test for differences between the population means for Case 1 (H_0: $\mu_1 - \mu_2 \leq \delta_0$). Modifications for Case 2 (H_0: $\mu_1 - \mu_2 \geq \delta_0$) are given in parentheses {}.

STEP 1: Calculate the sample mean $\overline{X}$ and the sample variance S_x^2 for sample 1 and compute the sample mean $\overline{Y}$ and the sample variance S_y^2 for sample 2.

STEP 2: Using Section 4.5, test whether the variances of the two populations are equal. If the variances of the two populations are not equal, compute:

$$S_{NE} = \sqrt{\frac{S_x^2}{m} + \frac{S_r^2}{n}}$$

If the variances of the two populations appear approximately equal, use Student's two-sample t-test (Section 3.3.1.1, Box 3.3-1).

STEP 3: Calculate:

$$t = \frac{\overline{X} - \overline{Y} - \delta_0}{S_{NE}}$$

Use Table A-1 of Appendix A to find the critical value t_{1-a} such that 100(1-a)% of the t-distribution with f degrees of freedom is below t_{1-a}, where

$$f = \frac{\left[\frac{S_X^2}{m} + \frac{S_Y^2}{n}\right]^2}{\frac{S_X^4}{m^2(m-1)} + \frac{S_Y^4}{n^2(n-1)}}$$

(Round f down to the nearest integer.)

If $t > t_{1-a}$ {$t < -t_{1-a}$}, the null hypothesis may be rejected. Go to STEP 5.

If $t \not> t_{1-a}$ {$t \not< -t_{1-a}$}, there is not enough evidence to reject the null hypothesis and therefore, the false negative error rate will need to be verified. Go to STEP 4.

STEP 4: If the null hypothesis (H_0) was not rejected, calculate either the power of the test or the sample size necessary to achieve the false positive and false negative error rates. To calculate the power of the test, assume that the true values for the mean and standard deviation are those obtained in the sample and use a statistical software package to generate the power curve of the two-sample t-test. A simple method to check on statistical power does not exist.

STEP 5: The results of the test could be:

1) the null hypothesis was rejected, and it seems $\mu_1 - \mu_2 > \delta_0$ {$\mu_1 - \mu_2 < \delta_0$};
2) the null hypothesis was not rejected, the false negative error rate was satisfied, and it seems $\mu_1 - \mu_2 \leq \delta_0$ {$\mu_1 - \mu_2 \geq \delta_0$}; or
3) the null hypothesis was not rejected, the false negative error rate was not satisfied, and it seems $\mu_1 - \mu_2 \leq \delta_0$ {$\mu_1 - \mu_2 \geq \delta_0$}, but this conclusion is uncertain because the sample size was too small.

Box 3.3-4: An Example of Satterthwaite's t-Test (Unequal Variances) for Simple and Systematic Random Samples

At a hazardous waste site, area 1 (cleaned using an in-situ methodology) was compared with a similar (but relatively uncontaminated) reference area, area 2. If the in-situ methodology worked, then area 1 should be approximately equal in average contaminant levels. If the methodology did not work, then area 1 should have a higher average than the reference area. Seven random samples were taken from area 1, and eight were taken from area 2. Because the contaminant concentrations in the two areas are supposedly equal, the null hypothesis is H_0: $\mu_1 - \mu_2 \leq 0$ (Case 1). The false positive error rate was set at 5% and the false negative error rate was set at 20% (β) if the difference between the areas is 2.5 ppb.

STEP 1:

	Sample Mean	Sample Variance
Area 1	9.2 ppm	1.3 ppm^2
Area 2	6.1 ppm	5.7 ppm^2

STEP 2: Using Section 4.5, it was determined that the variances of the two populations were not equal, and therefore using Satterthwaite's method is appropriate:

$$S_{NE} = \sqrt{1.3/7 + 5.7/8} = 0.9477$$

STEP 3:

$$t = \frac{9.2 - 6.1 - 0}{0.9477} = 3.271$$

Table A-1 was used with f degrees of freedom, where

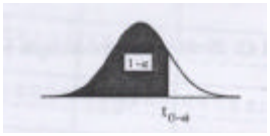
$$f = \frac{[1.3/7 + 5.7/8]^2}{\frac{1.3^2}{7^2(7-1)} + \frac{5.7^2}{8^2(8-1)}} = 10.307 \text{ (i.e., 10 degrees of freedom)}$$

(recall that f is rounded down to the nearest integer), to find $t_{1-a} = 1.812$.

Because $t > t_{0.95}$ (3.271 > 1.812), the null hypothesis may be rejected.

STEP 4: Because the null hypothesis was rejected, it would appear there is a difference between the two areas (area 1 being more contaminated than area 2, the reference area) and that the in-situ methodology has not worked as intended.

Table A-1: Critical Values of Student's t Distribution



Degrees of Freedom	1 - a								
	.70	.75	.80	.85	.90	.95	.975	.99	.995
1	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.536	0.691	0.866	1.074	1.340	1.753	2.131	2.602	2.947
16	0.535	0.690	0.865	1.071	1.331	1.746	2.120	2.583	2.921
17	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.527	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660
120	0.526	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576

Note: The last row of the table (∞ degrees of freedom) gives the critical values for a standard normal distribution (z), e.g., $t_{\infty,0.95}=z_{0.95}=1.645$